

Smith-Waterman Search

The *Smith-Waterman Search* plugin adds a complete implementation of the Smith-Waterman algorithm to UGENE.

To use the plugin open a nucleotide or protein sequence in the *Sequence View* and select the *Analyze Find pattern [Smith-Waterman]* item in the context menu. The *Smith-Waterman Search* dialog appears:

Smith-Waterman Search

Smith-Waterman parameters | Input and output

Enter pattern here

Search in
 Sequence
 Translation

Strand
 Both
 Direct
 Complement

Region
Whole sequence
1 - 199950

Smith-Waterman algorithm parameters

Algorithm version
Classic 2
Advanced..

Scoring matrix
dna
View..

Gap scores
Gap open: -10
Gap extension: -1

Results filtering strategy
Report results: filter-int
Minimal score: 90%

Help | Cancel | Search

First of all you need to specify the pattern to search for. The rest parameters are optional:

Search in — select either to search in the sequence or in its translation.

Strand — select the strand to search in: direct, complementary or both strands.

Region — specifies the region of the sequence that will be used to search for the pattern. By default, if a subsequence has been selected when the dialog has been opened, then the selected subsequence is searched for the pattern. Otherwise, the whole sequence is used. You can also input a custom range.

Algorithm version — version of the algorithm implementation. Non-classic versions produce the same results as classic but much faster. To use these optimizations our system must support these capabilities.

- Classic 2
- SSE2
- CUDA
- OPENCL

Scoring matrix — can be chosen from a bunch of matrices supplied with UGENE. To view a matrix selected click the *View* button.

Gap open — penalty for opening a gap.

Gap extension — penalty for extending a gap

Report results — simple heuristic which allows to filter intersected hits. If it is set to *none*, the algorithm may report large set of almost identical results in the same region.

Minimal score — another simple heuristic which measures sequences similarity. It is more convenient than using some abstract scores. If set to 100%, the algorithm will search for exact substring match.

The results of the search are saved as annotations or as multiple alignment. To set the saving parameters go to the *Input and output* tab of the dialog.

If you want to save the results as annotations input *the annotations saving* parameters (*Annotation name*, *Group name*, *Annotation type*, *Description* and a file to save the annotation to). Also you can add qualifier with corresponding pattern subsequences to result annotations. Check the corresponding checkbox for it.

If you want to save the results as multiple alignment select the following parameters:

The screenshot shows a dialog box titled "Smith-Waterman Search" with two tabs: "Smith-Waterman parameters" and "Input and output". The "Input and output" tab is active. It contains the following fields and controls:

- Save results as:** A dropdown menu set to "Annotations".
- Save annotation(s) to:** A section with three radio buttons:
 - Existing table: followed by a text field and a file selection icon.
 - Create new table: followed by a text field containing "/home/yalgaer/MyDocument.gb" and a file selection icon.
 - Use auto-annotations table
- Annotation parameters:** A section with four text fields:
 - Group name:** contains "<auto>" and has a refresh icon.
 - Annotation type:** a dropdown menu set to "Misc. Feature".
 - Annotation name:** contains "by type".
 - Description:** an empty text field.
- Add qualifier with corresponding pattern subsequences to result annotations

At the bottom of the dialog are three buttons: "Help", "Cancel", and "Search".

Here you can select a file to save the alignment to (*Alignment files directory path* parameter).

Using the *Set advanced options* checkbox you can select the saving options.

You can set the different templates for files names: create your own or create by using the following: [E] — adds a subsequence end position, [hms] — adds a time, [MDY] — adds a date, [S] — adds a subsequence start position, [L] — adds a subsequence length, [SN] — adds a reference sequence name prefix, [PN] — adds a pattern sequence name prefix, [C] — adds a counter.

You can create templates for alignment files names, reference subsequence names, pattern subsequence names and for pattern sequence name:

File Actions Settings Tools Window Help

Project

Name filter

Objects

- human_T1.fa
- [s] human_T1 (UCSC April 2002...
- P1_human_T1_1.aln**
 - [m] P1_human_T1_1

Consensus

Reference subsequence

Pattern subsequence

Alignment

	A	C	G	T	A	C	G	-
1	A	C	G	T	A	C	G	T
7	A	C	G	T	A	C	G	-

Ln 1/2 Col 1/8 Pos 1/8